CSIRO PUBLISHING

www.publish.csiro.au/journals/ijwf *International Journal of Wildland Fire*, 2004, **13**, 119–129

Spatial models for estimating fuel loads in the Black Hills, South Dakota, USA

Robin M. Reich[A,C], *John E. Lundquist*[B] *and Vanessa A. Bravo*[A,D]

[A]Department of Forest, Rangeland, and Watershed Stewardship, Colorado State University, Fort Collins, CO 80523, USA.
[B]Rocky Mountain Research Station, USDA Forest Service, Fort Collins, CO 80521, USA.
Telephone: +1 970 498 1095; fax: +1 970 498 1314; email: jlundquist@fs.fed.us
[C]Corresponding author. Telephone: +1 970 491 6980; fax: +1 970 491 6754; email: robin@cnr.colostate.edu
[D]Telephone: +1 970 491 6980; fax: +1 970 491 6754; email: vbravo@cnr.colostate.edu

Abstract. Fire suppression has increased fuel loadings and fuel continuity in many forested ecosystems, resulting in forest structures that are vulnerable to catastrophic fire. This paper describes the statistical properties of models developed to describe the spatial variability in forest fuels on the Black Hills National Forest, South Dakota. Forest fuel loadings (tonnes/ha) are modeled to a 30 m resolution using a combination of trend surface models to describe the coarse-scale variability in forest fuel, and binary regression trees to describe the fine-scale variability associated with site-specific variability in forest fuels. Independent variables used in the models included various Landsat TM bands, forest class, elevation, slope, and aspect. The models accounted for 55% to 72% of the variability in forest fuels. In spite of having highly skewed distributions, cross-validation showed the models to have nominal prediction bias. This paper also evaluates the feasibility of using the estimation error variance to explain estimation uncertainty. The models are allowing us to study the influence of small-scale disturbances on forest fuel loadings and diversity of resident and migratory birds on the Black Hills National Forest.

Additional keywords: binary regression trees; cross-validation; fuels; fuel loading; fuel variability; Landsat imagery.

Introduction

Forest managers have utilized models to aid in predicting fire behavior and to map varying fire spread scenarios during a given time period. FARSITE (Finney 1998), BEHAVE (Andrews 1986) and other fire behavior and growth decision support systems are based on a mathematical model for quantifying fire spread in surface fuels developed by Rothermel (1972). This model is composed of a series of calculations for heat required for ignition, propagating flux, reaction intensity, and effect of wind and slope. Weather and fuel parameters are used as input for these calculations. Spatial properties of the composition, quantity, size, compactness, and arrangement of fuels are major factors that determine initiation, intensity, and spread of forest wildfires (Pyne *et al*. 1996). Wildfire fuels are commonly split into three classes: aerial fuels, ground fuels, and surface fuels. Of these, surface fuels have the greatest influence on fire behavior. Surface fuels include trees less than 1.8 m tall, shrubs, grasses and forbs, litter, and coarse woody debris (Pyne *et al*. 1996). Small-scale disturbances alter abundance and composition of the standing dead tree and coarse woody debris components of surface fuels (Lundquist and Beatty 2002).

A fuel model is a 'stylized and simplified description of fuel for a mathematical fire behavior model' (Pyne *et al*. 1996), which includes a profile of characteristics used as input to fire models. According to Anderson (1982), the most important fuel characteristics are fuel load by size class; mean size and shape of each size class; compactness or bulk density; horizontal continuity; vertical arrangement; moisture content; and chemical content. Fuel load and size distribution and compactness determine fire sustainability; spatial continuity determines fire spread; and vertical pattern determines potential for fire to spread into the overstory canopy (Anderson 1982). The varying characteristics of fuels make them complex and diverse components of the ecosystem (Burgan 1987). Comprehensive fuel models take considerable sampling effort, and are largely impossible for developing spatial models based on ground surveys. To simplify and standardize fuel assessments, 13 standardized fuel models were established (Albini 1976). Because these fuel models are not correlated with vegetation type and actual conditions, estimates based on them commonly result in significant errors.

 10.1071/WF02049 1049-8001/04/010119

Because of the difficulty and expense in developing the required layers for comprehensive models, remote sensing has commonly been used to generate data for these models (Chuvieco and Salvas 1996). Typically, satellite imagery or aerial photographs have been used to map vegetation characteristics, and then assign fuel models to the various vegetation classes (Kourtz 1977; Miller and Johnston 1985; Wilson *et al*. 1994; Mark *et al*. 1995). The disadvantage of this approach is that the various components of forest structure are not always correlated with existing vegetation characteristics.

The use of remote sensing coupled with spatial analysis has been successful in detecting, predicting, and modeling total living and dead biomass in grasslands and shrublands (Friedl *et al*. 1994; Millington *et al*. 1994). Such approaches have had limited success in forest ecosystems because the ground is often obscured from view by the forest canopy, thereby making it difficult to discriminate among different types and amounts of fuel on the ground. Satellite imagery has also been used to study changes in landscape patterns caused by wildfires in several different forest types (Racine *et al*. 1985; Syrjanen *et al*. 1994; Turner *et al*. 1994; Razafimpanilo *et al*. 1995). Other studies have examined the application of GIS in predicting spatial spread of fires (Green *et al*. 1995; Perry *et al*. 1999; Keane *et al*. 2001). Several studies have examined the use of satellite images to develop fuel models (Rabii 1979; Shasby *et al*. 1981; Salazar 1982; Agee and Pickford 1985; Miller and Johnston 1985; Burgan *et al*. 1998; Roberts *et al*. 1998).

The only reliable method of estimating forest fuels is to collect intensive field measurements, which are time consuming and lack adequate replication. Such data are generally summarized on a stand level in a GIS representing vegetation types. Because these scales are too coarse for fine-scale fire spread models, they may produce unrealistic results. In some instances, field data have been spatially interpolated to generate locally oriented GIS-based fire danger models that cover small areas at a high spatial resolution (Chuvieco and Salvas 1996). Keane *et al*. (2001) provides an excellent review and discussion of past, present, and future approaches for modeling forest fuels for fire management at multiple scales.

We recently developed a spatial model that used satellite imagery and field data to predict diversity of migratory birds in the Black Hills. This method enabled the generation of spatial models that predicted bird diversity over a broad range of scales, from sub-stands to multiple watersheds. In the study described below, we adapt this method to develop models describing the fine-scale spatial variability of forest fuels over the entire Black Hills National Forest, South Dakota. This study is part of an examination into the relative influences of diseases on fuel loading and their relative importance compared to other co-occurring disturbances. Our aim here was to develop a reliable tool for predicting the distribution of fuel loadings across both fine and broad spatial scales using remotely sensed data, and spatial models.

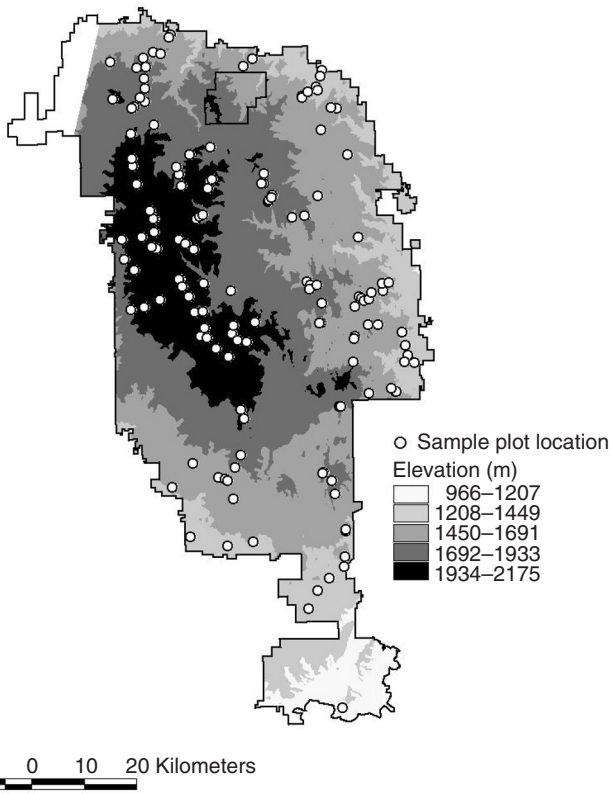

Fig. 1. Location of 151 sample plots overlaid on a digital elevation model of the Black Hills National Forest.

Our approach incorporates new statistical techniques, which are cross-validated to provide error estimates and classification accuracies. This is an important aspect of this paper in that it is important to be able to understand the type of errors associated with models being used in any future analysis.

Methods

Study site

This study was conducted in the Black Hills National Forest in west central South Dakota, USA. The Black Hills are the easternmost outlier of the Rocky Mountains and cover ~650 000 ha. Elevation reaches 2300 m (Froiland 1990). Forests occur between 966 m and 2175 m (Fig. 1). Precipitation averages ~750 mm per year, mostly as rain during summer months. *Pinus ponderosa* Douglas ex P. Laws is the dominant tree species. At the higher elevations, *Picea glauca* (Moench) Voss is the dominant tree species, while both *Populus tremuloides* Michx. and *Quercus macrocarpa* Michx. (Hoffman and Alexander 1987) dominate in the northern part of the forest.

Sample plots

A total of 151 sample plots each 30 m × 30 m were randomly located throughout the Black Hills National Forest using a stratified design (Fig. 1). Using an existing vegetation map,

the forest was stratified into broad forest classes (pine, spruce, oak, aspen, and meadows). Using an initial sample of 50 plots, the variability in each stratum was determined. This information was used to allocate the remaining sample plots proportional to the variability observed in forest structure and fuel loadings. Sample plots were oriented in a north–south direction and georeferenced using GPS. A 42 m transect was established diagonally across the sample plots from the south-west corner to the north-east corner. Three 1 m transects were established 7 m, 21 m, and 35 m along the 42 m transect. Data collected on the 30 m × 30 m sample plots included average tree height (m), height to the base of the live crown (m), canopy closure (%), average height (m), diameter (m), number of shrubs per ha, and total tree basal area (m^2/ha). The information on shrubs was used to estimate shrub volumes per ha, which included air space.

Estimation of fuel loading

Using procedures developed by Brown *et al*. (1982), we used planar intersect sampling to estimate fuel loadings in tonnes/ha for the following size classes: <0.6 cm; 0.6–2.5 cm; 2.5–7.6 cm; and >7.6 cm sound and dead. Fuel loadings in size classes under 7.6 cm were estimated by counting the number of intercepts on the three 1 m transects by species group (pine, spruce, aspen hardwood). Fuel loadings for woody material greater than 7.6 cm were estimated by measuring the diameter of all intercepts on the 42 m transect, by species group. In addition to the intercept data, information was collected on the depth (cm) of the litter and duff on the three 1 m transects, along with estimates of the height of the woody fuels (cm) above the forest floor. All fuel loading estimates were adjusted for slope and species group.

GIS and Landsat TM data

We derived GIS grids of elevation, slope, and aspect from digital elevation models using ArcView® (ESRI 1998) to a 30 m spatial resolution, corresponding to the resolution of the field data. Grids of spectral bands 1–5, 6l, 6h, 7 and 8 of a cloud-free, 2001 Landsat TM image corresponding to the timing of the field data collection was also created. Spectral bands 6l and 6h were thermal bands (60 m resolution), while band 8 was a panchromatic image (5 m resolution). These latter three bands were resampled [Resample function, nearest neighbor, Grid Module (ARC/INFO®)] (ESRI 1995) to a 30 m spatial resolution, corresponding to the resolution of the field data. A grid of six vegetation classes, modeled to a 30 m spatial resolution, was derived from an independent dataset using a binary classification tree (Breiman *et al*. 1984). Individual class accuracy of the vegetation model, were: pine, 97%; pine/deciduous, 96%; spruce, 90%; riparian, 95%; aspen/deciduous, 93%; meadows/grass, 87%. Values for all grid layers of information were derived for each 30 m sample plot using Avenue (ESRI 1998) code.

Modeling forest fuels

Modeling of forest fuels was accomplished in two stages. In the first stage, multiple regression analysis (OLS: Reich and Davis 1998) was used to explore the coarse-scale variability in forest fuels as a function of elevation, slope, aspect, and Landsat TM bands. To account for differences among forest classes, dummy variables were introduced in the models as interactions with elevation, slope, aspect, landform, and the eight Landsat bands. For each component of forest fuels modeled, we used a stepwise procedure to identify the best subset of independent variables to include in the regression models.

In the second stage, we modeled the error (i.e. residuals) associated with the regression models using binary regression trees. A binary regression tree is a non-parametric approach to regression that compares all possible splits among the independent (continuous) variables using a binary partitioning algorithm that maximizes the dissimilarities among groups. Once the algorithm partitions the data into new subsets, new relationships are developed, assessed, and split into new subsets. The algorithm recursively splits the data in each subset until either the subset is homogeneous or the subset contains too few observations (e.g. <5) to be split further. Interpolation using RTs is relatively insensitive to sparse data. Independent variables considered in the RT included elevation, slope, aspect, landform, Landsat TM bands, and forest class, the latter being treated as a categorical variable. To avoid over-fitting the models, a 10-fold cross-validation procedure (Efron and Tibshirani 1993) was used to identify the tree size that minimizes the total deviance associated with the trees.

Grids representing the various forest fuel components were generated for the best-fitting regression model using the model's parameter estimates. Similarly, grids representing the error in each regression model were generated by passing each grid for the appropriate independent variable through the regression trees. The final predicted surfaces of the different components forest fuels were obtained from the sum of the two grids.

Semi-variograms were used to evaluate spatial dependencies among the residuals from the various forest fuel models. If the residuals exhibited spatial dependencies, a spatial autoregressive model was used to obtain generalized least squares estimates of the regression coefficients associated with the TS model (Upton and Fingleton 1985). The model residuals were reevaluated to ensure the removal of the spatial dependencies. In fitting the spatial autoregressive models, a spatial weight matrix (i.e. a block diagonal matrix) based on inverse distance weighting was used to represent the spatial dependencies among the sample plots.

The effectiveness of the final models was evaluated using a goodness-of-prediction statistic (G) (Agterberg 1984; Kravchenko and Bullock 1999; Guisan and Zimmermann 2000; Schloeder *et al*. 2001). The G-value measures how effective a prediction might be relative to that which could

have been derived by using the sample mean (Agterberg 1984):

$$G = \left(1 - \left\{\sum_{i=1}^{n}\left[Z_i - \hat{Z}_i\right]^2 \Big/ \sum_{i=1}^{n}\left[Z_i - \overline{Z}\right]^2\right\}\right), \quad (1)$$

where Z_i is the observed value of the ith observation, $\hat{Z}_i$ is the predicted value of the ith observation, and $\overline{Z}$ is the sample mean. A G-value equal to 1 indicates perfect prediction, a positive value indicates a more reliable model than if one had used the sample mean, a negative value indicates a less reliable model than if one had used the sample mean, and a value of zero indicates that the sample mean should be used to estimate Z.

Model evaluation

A 10-fold cross-validation (Efron and Tibshirani 1993) was used to estimate the prediction error for each forest fuel component. The data were split into $K = 10$ parts consisting of ~15 sample plots. For each part, the models were fitted to the remaining $K - 1 = 9$ parts of the data. The fitted model was used to predict the part of the data removed from the modeling process. This process was repeated 10 times so that each sample plot was excluded from the model fitting step and its response predicted. The prediction errors can then be inferred from the predicted minus actual values. Repeating this process over many deleted subsets allows an assessment of the variability of prediction error. While it may be desirable to assess the uncertainty in the models using an independent dataset, this may not always be feasible because of time and cost constraints. The cross-validation procedures used in this paper have become a popular method of assessing accuracy and prediction since the articles by Stone (1974) and Geisser (1975).

To evaluate the effectiveness of the models, we computed various measures of prediction error. Prediction bias (Williams 1997) was calculated for each validation dataset as a percentage of the true value. Accuracy (Kravchenko and Bullock 1999; Schloeder *et al*. 2001) was measured by the mean absolute error (MAE), which is a measure of the sum of absolute residuals (i.e. actual minus predicted) and the root mean squared error (RMSE), which is the square root of the sum of squared residuals. Small MAE values indicate a model with few errors, while small values of RMSE indicate more accurate predictions on a point-by-point basis (Schloeder *et al*. 2001). To assess the estimation uncertainty in the models (Isaaks and Srivastava 1989), we calculated the estimation error variance (EEV), $\hat{\sigma}_i^{2*}$ for each observation in the dataset:

$$\hat{\sigma}_i^{2*} = MSE\left[(X_i^*)'(X'X)(X_i^*) + 1\right] + 2MSE(\mathrm{RT}) + 2Cov(\hat{Y}, \hat{\eta}), \quad (2)$$

where *MSE* is the regression mean squared error for the TS model fitted using $K-1 = 9$ parts of the data, X is a matrix of independent variables used in fitting the model, X_i^* is a vector of independent observations not used in fitting the model, *MSE*(RT) is the mean squared error of the RT used to describe the error in the TS model, and $Cov(\hat{Y}, \hat{\eta})$ is the covariance between the estimated values ($\hat{Y}$), from the TS model and the predicted residuals ($\hat{\eta}$) . The consistency between the EEV and the observed estimation errors (i.e. true errors), $e_i^* = (Z_i - Z_i^*)$, was calculated using the standard mean squared error (SMSE) (Hevesi *et al*. 1992):

$$\mathrm{SMSE} = \frac{1}{n}\sum_{i=1}^{n}\frac{(e_i^*)^2}{\hat{\sigma}_i^{2*}}. \quad (3)$$

EEVs were assumed consistent with true errors if the SMSE fell within the interval $[1 \pm 2(2/n)^{-1/2}]$ (Hevesi *et al*. 1992). Paired t-tests ($\alpha = 0.05$) were used to test for differences between the mean estimation errors and zero. The EEVs were also used to construct 95% confidence intervals around individual estimates. Coverage rates were calculated as the proportion of individual confidence intervals that contained the true value.

Custom fuel models

A k-means clustering algorithm was used to partition the sample data of fuel loadings into three groups to minimize the within-group sum of squares (Hartigan and Wong 1979). Variables used in the analysis included the <0.6 cm; 0.6–2.5 cm; 2.5–7.6 cm fuel loadings, fuel height and shrub volume. The first three variables are important components required to define custom fuel models in FARSITE. The clustering algorithm was then used to generate a GIS grid of the Black Hills where each forested pixel was assigned to one of three fuel classes. We did this as a demonstration of how the various GIS grids developed in this paper could be used in describing the spatial distribution of fuel loadings and are not intended to represent valid fuel models.

Results

Modeling forest structure

Summary statistics for fuel loading data used in developing the models are provided in Table 1. The majority of the sample distributions were highly skewed to the right, which influenced the final fit of the models. It was not possible to transform the data to sufficiently remove this skewness. However, residuals plots, and plots of predicted *v.* observed fuel values did not show any trends to suggest a systematic bias in any of the models.

Table 2 summarizes the independent variables used to describe the coarse-scale variability in fuel loadings. Fuel loadings were linearly correlated with the topographic and Landsat TM data, and these linear relationships varied significantly among forest type. For example, classes of fuel loadings observed in spruce and aspen stands varied significantly from those observed in pine stands. All of the

Table 1. Summary statistics of observed and estimated forest fuels for the Black Hills National Forest from 10-fold cross-validation

Statistic	Fuel height (cm)		Litter (cm)		Duff (cm)		0–0.6 cm (t/ha)	
	Fuel sampled data	Modeled estimates	Fuel sampled data	Modeled estimates	Fuel sampled data	Modeled estimates	Fuel sampled data	Modeled estimates
N	151	151	151	151	151	151	151	151
Mean	5.84	6.43	3.53	3.73	2.23	2.23	0.32	0.36
s.d.	9.52	7.39	2.54	1.98	1.75	1.45	0.47	0.38
CV%	162.7	115.1	72.0	53.2	78.7	64.3	153.2	108.5
Minimum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
First quantile	0.84	1.78	2.11	2.44	0.84	1.19	0.02	0.07
Median	2.79	3.33	3.38	3.69	1.70	2.01	0.16	0.22
Third quantile	5.61	9.32	4.65	4.65	3.38	3.38	0.36	0.49
Maximum	64.33	54.66	23.39	11.91	9.32	5.59	2.46	2.01
Bias%		−9.68		−5.54		0.0		−13.57
	0.6–2.5 cm (t/ha)		2.5–7.6 cm (t/ha)		<7.6 cm (t/ha)		>7.6 cm (t/ha)	
N	151	151	151	151	151	151	151	151
Mean	3.50	4.06	14.28	18.18	17.95	19.05	14.05	14.32
s.d.	6.14	5.09	23.16	29.32	28.15	20.17	15.44	11.39
CV%	177.2	125.5	163.3	161.4	156.7	108.7	109.9	79.6
Minimum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
First quantile	0.0	0.83	0.0	2.42	0.83	4.55	1.30	5.54
Median	1.43	3.27	5.33	8.00	7.82	11.07	9.06	12.69
Third quantile	3.97	4.75	17.04	21.36	18.91	25.98	20.55	21.25
Maximum	43.17	49.41	123.36	187.14	169.00	92.54	66.64	51.60
Bias%		−16.02		−27.27		−6.04		−1.87

Table 2. Description of the multiple regression models for describing the coarse-scale variability in forest fuels on the Black Hills National Forest

Circles indicate the inclusion of a variable as a main effect, squares indicate an interaction between forest classes and topographic and Landsat TM bands. Forest classes were treated as dummy variables with pine being the default forest type

Model	Topography[A]			Landsat TM bands[B]									Forest class[C]				
	E	S	A	B1	B2	B3	B4	B5	B6L	B6H	B7	B8	P	R	SP	AS	M
Duff (cm)	●		●■	●		●	■	■		■			●	●	■	■	■
Litter (cm)	■	■	■	■		■					●		●	■	■	■	■
Fuel height (cm)	●	■	■	■			■				■		●	■	■		
0–0.6 cm (t/ha)		■					■						●		■	■	
0.6–2.5 cm (t/ha)	■	■	■	■	■	■		■					●		■	■	
2.5–7.6 cm (t/ha)				●			■	■			■		●		■	■	
<7.6 cm (t/ha)	■	■	■				■	■	●		■		●		■	■	
>7.6 cm (t/ha)	■	●■	■		■●						●		●	■	■	■	■

[A]E, elevation; S, slope; A, aspect.
[B]B1, band 1; B2, band 2; B3, band 4; B5, band 5; B6L, band 6 low; B6H, band 6 high; B7, band 7; B8, band 8.
[C]P, pine; R, riparian; SP, spruce; AS, aspen/deciduous; M, meadows/openings.

topographic, Landsat TM, and forest class variables (Table 3) were also used in the regression trees to describe the error in one or more of the regression models. Notably, however, none of the forest classes contributed to classifying residuals in the models for the depth of the duff and litter, indicating that the regression model for these components accounted for all the structural variability due to species composition. Forest classes were important in nearly all other models of fuel loadings. The tree sizes selected to minimize the total deviance in the regression trees ranged from 23 to 49 splits.

The overall contribution of the models (Table 4) in describing forest fuels varied with the model. The regression models alone explained 34% (<0.6 cm fuels) to 45% (>7.6 cm fuels) of the observed variability in fuel loadings. The regression trees accounted for an additional 19% (0.6 cm–2.5 cm fuels) to 35% (<7.6 cm fuels) of the unexplained variability in the TS models. Overall model performance ranged from a low of 0.55 for <0.6 cm fuels to a high of 0.71 for 2.5–7.6 cm fuels. The remaining six models had G-values ranging from 0.61 to 0.69. The final models are displayed in Fig. 2.

Table 3. Variables (circles) used in the binary regression trees to describe the error in the trend surface model of components of forest fuels on the Black Hills National Forest

Forest type was treated as a categorical variable in the regression trees

Model	Topography[A]			Landsat TM bands[B]									Forest class[C]				
	E	S	A	B1	B2	B3	B4	B5	B6L	B6H	B7	B8	P	R	SP	AS	M
Duff (cm)	•	•	•	•			•	•	•	•		•					
Litter (cm)	•	•	•			•	•	•	•	•	•						
Fuel height (cm)			•	•		•			•	•	•		•	•	•	•	•
0–0.6 cm (t/ha)	•		•	•	•	•			•	•	•	•	•		•	•	
0.6–2.5 cm (t/ha)	•	•	•				•	•	•				•	•	•	•	•
2.5–7.6 cm (t/ha)	•	•	•		•	•	•			•		•	•	•	•	•	•
< 7.6 cm (t/ha)	•	•	•	•		•	•		•	•	•		•	•	•	•	•
> 7.6 cm (t/ha)	•		•	•		•	•		•	•		•	•	•	•	•	•

[A]E, elevation; S, slope; A, aspect.
[B]B1, band 1; B2, band 2; B#, band 4; B5, band 5; B6L, band 6 low; B6H, band 6 high; B7, band 7; B*, band 8.
[C]P, pine; R, riparian; SP, spruce; AS, aspen/deciduous; M, meadows/openings.

Table 4. Overall model performance (G-statistic) of the multiple regression models (TS) and the use of binary regression trees (RT) in describing the errors in the multiple regression models of forest fuels on the Black Hills National Forest

Model	TS	RT	TS + RT
Fuel height (cm)	0.355	0.261	0.616
Litter (cm)	0.348	0.264	0.612
Duff (cm)	0.345	0.324	0.669
0–0.6 cm (t/ha)	0.338	0.210	0.548
0.6–2.5 cm (t/ha)	0.417	0.194	0.611
2.5–7.6 cm (t/ha)	0.431	0.282	0.713
Small (<7.6 cm) (t/ha)	0.346	0.349	0.695
Large (>7.6 cm) (t/ha)	0.447	0.221	0.668

Looking at these figures one can see that some fuel classes (i.e. 0.6–2.5 cm, 2.5–7.6 cm, and fuel height) have similar spatial patterns. These patterns follow trends in elevation and species distribution.

Model evaluation

Prediction bias was nominal (Table 5) for all models. Minimum, maximum, and quartile values showed that estimated and observed value distributions were similar for all models. Estimation errors for the depth of the duff and litter had a similar spread in the estimation errors. In terms of the fuel models, the 2.5–7.6 cm, <7.6 cm and >7.6 cm fuel models had the largest spread in terms of the estimation errors, while the <0.6 cm fuel model had the least spread. The mean estimation errors did not differ significantly from zero (P-value ≥ 0.050). The MAE was smaller than the RMSE for all models indicating that, in general, the models are more accurate in predicting regional or global means than on a point-by-point basis.

SMSE results (Table 5) showed that the computed EEVs were statistically consistent with the true errors for all models, except for the 2.5–7.6 cm fuels, as they were within the interval [0.7698–1.2302] (Hevesi *et al*. 1992). This suggests that EEVs could be used to assess estimates of uncertainty for new observations. The EEV for the 2.5–7.6 cm fuels was lower than the true error (SMSE > 1) and, therefore, underestimated the true error. The 0.95 confidence coverage rates ranged from a low of 0.90 for the 2.5–7.6 cm fuel model to a high of 0.99 for the litter model. Half of the model had coverage rates less than 0.95. This suggests that confidence intervals constructed using the EEV may not be large enough to insure a 95% confidence interval around our estimates.

Custom fuel models

To better reflect fire behavior, models such as FARSITE allow the user to define custom fuel models. Important components required to define such models include the <0.6 cm, 0.6–2.5 cm, and 2.5–7.6 cm fuel loadings, which were modeled in this paper. Using these variables and information on shrub volumes and fuel height, we identified three fuel classes. Table 6 summarizes the within-group statistics of the variables used in the clustering. The fuel classes were number 20, 21 and 22, respectively. Figure 3 depicts the spatial distribution of the three fuel classes for a 2700 ha portion of the Black Hills National Forest. Fuel class 21 covers ~77% of the forested area and is characterized by relatively low fuels. Fuel class 20 covers ~19% of the forested area tends to have a higher fuel loading in the 2.5–7.6 cm class and slightly higher volume of shrubs as compared to fuel class 21. Fuel class 22 covers only 4% of the area, but tends to have the highest amounts of fuel. To be able to use these fuel classes in the program FARSITE, we would also need information on the area-to-volume ratio of fuels, which was not collected at the time of this study. This information could have been collected, along with other required information and modeled using the procedures presented in this paper. One could also develop GIS grids for canopy closure (required by FARSITE), average tree height, height to the base of the live crown and use this information as input to FARSITE to better describe forest conditions.

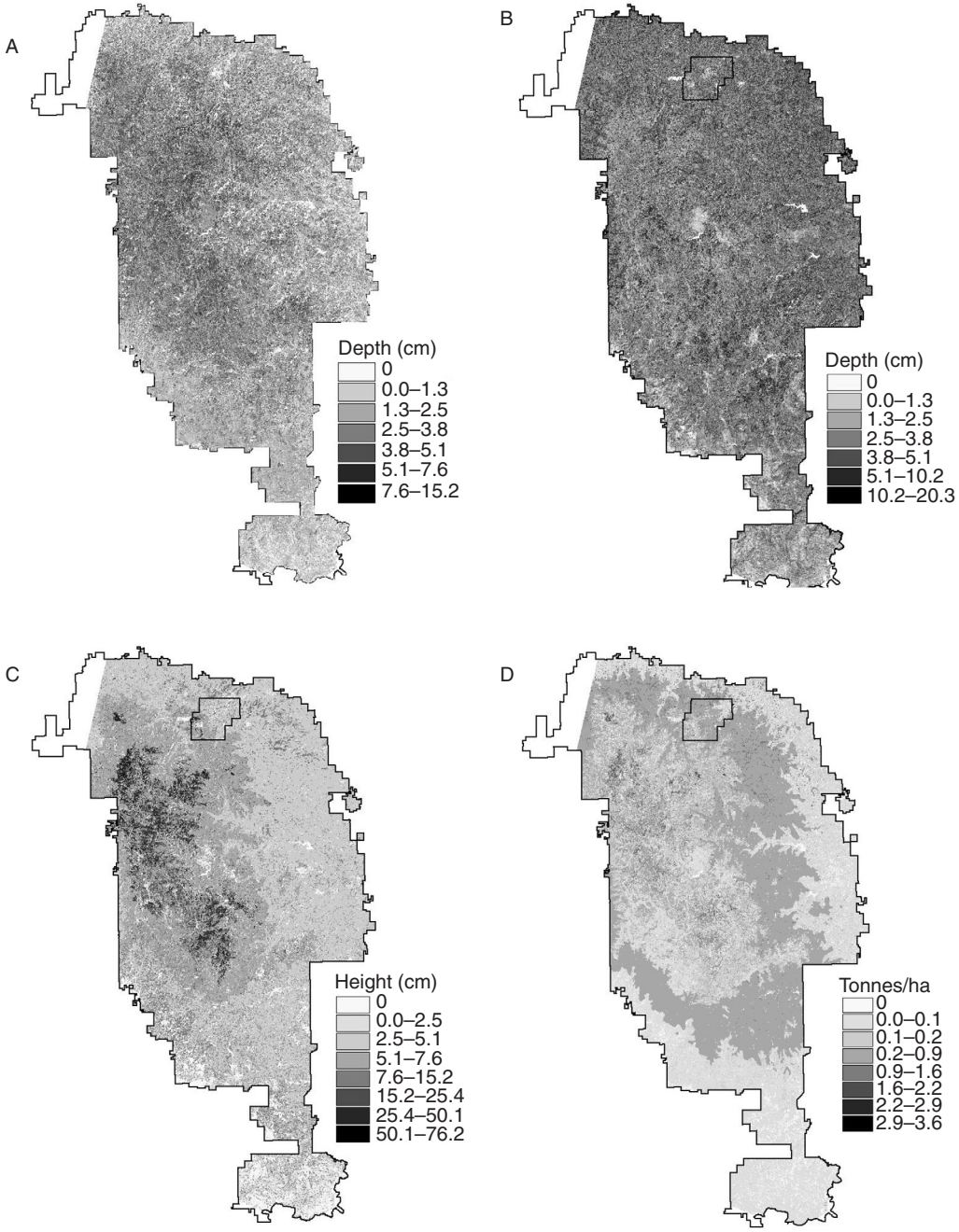

Fig. 2. Spatial distribution of (*A*) depth of duff (cm), (*B*) depth of litter (cm), and (*C*) fuel height (cm), and forest fuel loadings (t/ha) for the following size classes (cm): (*D*) <0.6, (*E*) 0.6–2.5, (*F*) 2.5–7.6, (*G*) <7.6, and (*H*) >7.6. Similarities in the spatial distribution of forest fuels (*C*, *E*, *F*, *G*) are due to a strong linear correlation with topographic variables such as elevation and slope, and the type of forest (pine, spruce, aspen).

Discussion

Our models of forest fuels can be very useful in that they provide detailed estimates of the structural components of forest fuels at a fine-scale (30 m resolution) with relatively high accuracy. These models, in general, produce greater accuracy and spatial resolution than traditional techniques based primarily on remotely sensed data, such as Landsat TM data, which are limited to what the sensors can directly detect. Our results are also superior to traditional methods of mapping forest fuels in that they are generally insufficient at predicting the spatial variability in forest fuels within a given stand.

Our models describe 55% (0–0.6 cm fuels) to 71% (2.5–7.6 cm fuels) of the spatial variability in the components of forest fuels observed on the field plots. The poor ability of some of the models to describe the various components

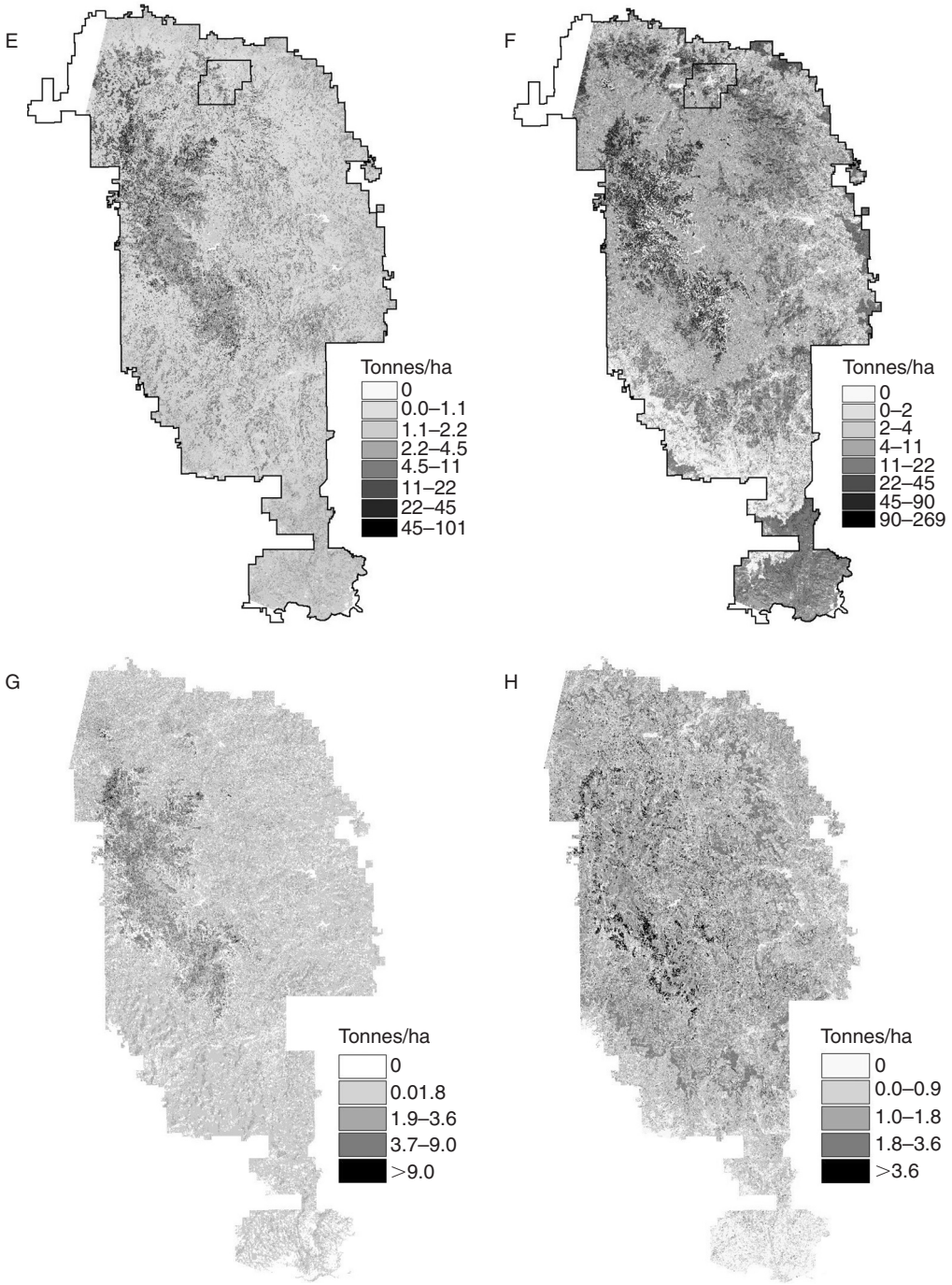

Fig. 2. (*Continued*)

of forest fuels may be due, in part, to certain types of management activities on the forest, such as controlled burning and thinning. The models assume that similar forest stands have similar characteristics with respect to forest fuels. It may be possible to include past management activities in the models to account for this variability and to improve the overall accuracy of the models. All models provided unbiased (P-value > 0.05) or marginally biased estimates of forest fuels. In the latter models, the bias was not large enough to limit the use of the models for predictive purposes.

Results of the 10-fold cross-validation indicated that the estimated error variances for the models provided statistically consistent estimate of the true prediction errors associated with the models. These results suggest that the estimated error variances could be used to assess estimates of uncertainty when the models are applied to non-sampled plot locations.

Table 5. Summary statistics of estimation errors of the forest fuels models for the Black Hills National Forest based on the 10-fold cross-validation

IQR = interquartile range; MAE = mean absolute error; RMSE = root mean square error; SMSE = standardized mean square error

Statistic	Fuel height (cm)	Litter (cm)	Duff (cm)	0–0.6 cm (t/ha)	0.6–2.5 cm (t/ha)	2.5–7.6 cm (t/ha)	<7.6 cm (t/ha)	>7.6 cm (t/ha)
N	151	151	151	151	151	151	151	151
Mean	−0.59	−0.20	0.0	−0.04	−0.56	−3.90	−1.10	−0.27
IQR	6.37	2.19	2.34	0.45	4.77	19.68	20.02	17.04
MAE	6.86	1.90	1.50	0.34	4.68	20.13	18.25	12.71
RMSE	11.40	2.87	1.98	0.51	8.16	36.36	28.56	17.64
SMSE	0.93	1.21	0.86	0.93	1.21	1.40	0.81	0.87
0.95 confidence coverage rate	0.93	0.99	0.95	0.93	0.95	0.90	0.95	0.93

Table 6. Summary statistics of variables used to identify three groups of forest fuels using a *k*-means clustering algorithm

Fuel model	Fuel height (cm)	0–2.5 cm fuels (t/ha)	0.6–2.5 cm fuels (t/ha)	2.5–7.6 cm fuels (t/ha)	Shrub volume (t/ha)
20	8.2	0.8	9.4	45.3	24.8
21	4.3	0.3	2.1	6.1	18.4
22	29.0	1.3	24.2	116.7	19.1

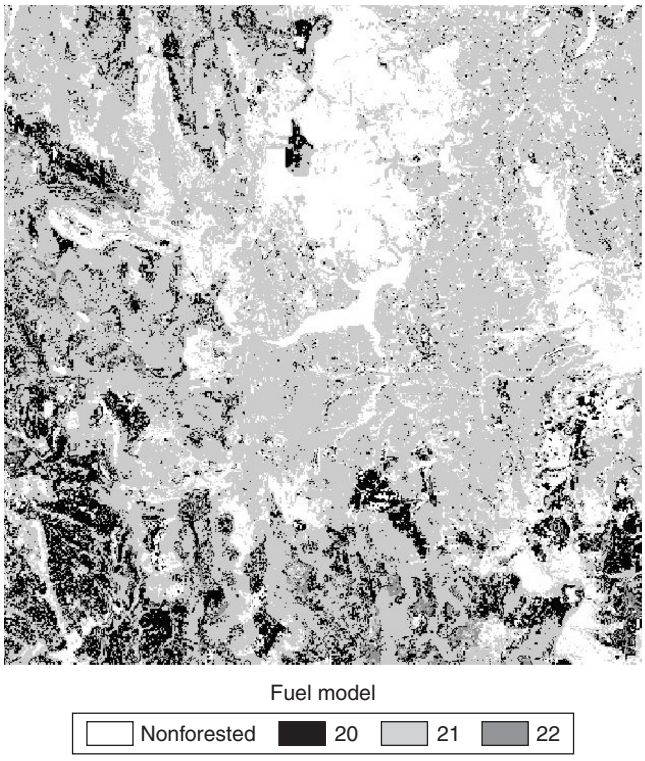

Fig. 3. Spatial distribution of three custom fuel classes for a 27 000 ha portion of the Black Hills National Forest. Fuel classes were based on the amount of 0–0.6 cm, 0.6–2.5 cm, and 2.5–7.6 cm fuels, fuel height (cm) and shrub volume (m^3/ha).

The ability to calculate estimation uncertainties allows us to develop GIS layers showing the computed estimation errors as well as place confidence intervals around our estimates. We assume that the estimation uncertainties are at their lower limit because the data used in this study were assumed to be error free. Other sources of errors that could have influenced the performance of the models include the sparseness of the field plots, errors in the forest classes and registrations errors with the field plots.

The models presented in this paper can be used to obtain estimates and associated standard errors or prediction for any specified geographical region (i.e. forest stand, management unit forest) within the Black Hills National Forest using the appropriate formula. This approach to estimating forest fuels is more cost effective than estimates based on field sampling alone. Another disadvantage of relying solely of field sampling is that there may, or may not be, sample plots available in every forest stand, especially if the stands are difficult to reach, small in size, or irregularly shaped.

Our initial interest in developing and evaluating these models was aimed at characterizing and quantifying the relative impacts of diseases and other small-scale disturbances on fuel loading. The characteristics of various fuel components depend to some extent on how fuel was formed (Lundquist and Beatty 2002). Spatial distributions in space and time probably mirror their causes, but determining causes can be difficult. Tree diseases and other small-scale disturbances can be major contributors to wildfire fuels, and certainly enhance spatial and temporal variability of fuels within and among forest stands. According to Keane *et al*. (2001), such variability is a major hurdle to generating accurate distribution maps for fuels.

Conclusions

This study presents a detailed description of a framework for mapping fuels based on spatially independent variables and

for characterizing the error inherent in the technique. Most wildfires are less than the size of managed stands. The most troublesome and destructive fires are much larger. Predicting spread and intensification of forest fires requires detailed information on the spatial distribution of forest fuels at a fine enough spatial resolution to accurately describe the dynamics of a fire and at a large enough scale to include potential routes of spread across the landscape. This study developed a series of models describing the spatial distribution of fuel loading on the Black Hills National Forest to a 30 m resolution by combining analyses of satellite images and field surveys. The models provide unbiased estimates of the various components of forest fuels as well as estimates of the prediction variance associated with individual estimates.

Good management requires decision support tools that are not only reliable, but also economical and timely. Fire spread models and other decision support systems are used to develop strategies about managing wildfires. Remote sensing enables a way of interpolating conditions at locations that are difficult or impossible to physically inspect. Although the statistical approach used to generate the spatial models would need adaptation to make them user friendly, these models should eventually be useful to managers that need predictions of fuel load distributions to make decisions about fire hazards and risks.

The low accuracy associated with existing approaches for modeling forest fuels is attributed to insufficient or low quality field data used in mapping vegetation types. A major limitation of fuels mapping directly from remotely sensed imagery is the inability for passive sensing instruments to penetrate the canopy. Also, existing techniques cannot provide estimates of uncertainty associated with their estimates. By modeling the field data as a function of topographic and remotely sensed imagery, it is possible to produce accurate estimates of forest fuels across vegetation types to a high degree of spatial resolution. It is also possible to model components of forest structure to the same spatial resolution as the fuel models. This would ensure that the layers required by a particular fire spread model all have the same spatial resolution thus ensuring consistent estimates.

References

Agee JK, Pickford SG (1985) 'Vegetation and fuel mapping of North Cascades National Park.' Final Report. (College of Forest Resources: Seattle) 111 pp.

Agterberg FP (1984) Trend surface analysis. In 'Spatial statistics and models'. (Eds GL Gaile and CJ Willmott) pp. 147–171. (Reidel: Dordrecht)

Albini FA (1976) 'Estimating wildfire behavior and effects.' USDA Forest Service, Intermountain Forest and Range Experiment Station General Technical Report INT-30. (Ogden, UT) 74 pp.

Anderson HE (1982) 'Aids to determining fuel models for estimating fire behavior.' USDA Forest Service Research Paper INT-305. 22 pp.

Andrews PL (1986) 'BEHAVE: fire prediction and modeling system: BURN subsystem, Part 1.' USDA Forest Service, Intermountain Forest and Range Experiment Station General Technical Report INT-194. (Ogden, UT) 130 pp.

Breiman L, Friedman JH, Olshen RA, Stone IJ (1984) 'Classification and regression trees.' (Wadsworth: Belmont, CA). 368 pp.

Brown JK, Oberheu RD, Johnston CM (1982) 'Handbook for inventorying surface fuels and biomass in the Interior West.' USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-129. 48 pp.

Burgan RE (1987) 'Concepts and interpreted examples in advanced fuel modeling.' USDA Forest Service, Intermountain Research Station General Technical Report INT-238. (Ogden, UT) 40 pp.

Burgan RE, Klaver RW, Klaver JM (1998) Fuel models and fire potential from satellite and surface observations. *International Journal of Wildland Fire* **8**, 159–170.

Chuvieco E, Salvas J (1996) Mapping the spatial distribution of forest fire danger using GIS. *International Journal of Geographical Information Systems* **10**, 333–345. doi:10.1080/026937996138070

Efron B, Tibshirani RJ (1993) 'An introduction to the bootstrap.' (Chapman and Hall: New York)

ESRI (1995) 'ARC/INFO® Software and on-line help manual.' (Environmental Research Institute, Inc.: Redlands, CA)

ESRI (1998) 'ArcView® 3.1.' (Environmental Research Institute, Inc.: Redlands, CA)

Finney MA (1998) 'FARSITE: Fire area simulator-area model. Development and evaluation.' USDA Forest Service, Rocky Mountain Research Station Report RMRS-RP-4. (Ogden, UT) 47 pp.

Friedl MA, Michaelson J, Davis FW, Walker H, Schimel DS (1994) Estimating grassland biomass and leaf area index using ground and satellite data. *International Journal of Remote Sensing* **15**, 1401–1420.

Froiland SG (1990) 'Natural history of the Black Hills and Badlands.' (The Center for Western Studies, Augustana College: Sioux Falls, SD) 225 pp.

Geisser S (1975) The predictive sample reuse method with applications. *Journal of the American Statistical Association* **70**, 320–328.

Green K, Finney M, Campbell J, Weinstein D, Landrum V (1995) Using GIS to predict fire behavior. *Journal of Forestry* **93**, 21–25.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* **135**, 147–186. doi:10.1016/S0304-3800(00)0354-9

Hartigan JA, Wong MA (1979) A *k*-means clustering algorithm. *Applied Statistics* **28**, 100–108.

Hevesi JA, Istok JD, Flint AL (1992) Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. *Journal of Applied Meteorology* **31**, 661–676. doi:10.1175/1520-0450(1992)031<0661:PEIMTU>2.0.CO;2

Hoffman GR, Alexander RR (1987) 'Forest vegetation of the Black Hills National Forest of South Dakota and Wyoming: A habitat type classification.' USDA Forest Service, Rocky Mountain Forest and Range Experiment Station Research Paper RM-276. (Fort Collins, CO) 48 pp.

Isaaks EH, Srivastava RM (1989) 'An introduction to applied geostatistics.' (Oxford University Press: New York) 561 pp.

Keane RE, Burgan R, van Wagtendonk J (2001) Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire* **10**, 301–319. doi:10.1071/WF01028

Kourtz PH (1977) An application of Landsat digital technology to forest fire fuel mapping. In 'Fire ecology in resource management: A workshop.' (Ed. DE Dube) pp. 79–81. Alberta. Information Report NOR-X-210. (Northern Forest Research Centre: Edmonton)

Kravchenko A, Bullock DG (1999) A comparative study of interpolation methods for mapping soil properties. *Agronomy Journal* **91**, 393–400.

Lundquist JE, Beatty JS (2002) A method for characterizing and mimicking forest canopy gaps caused by different disturbances. *Forest Science* **48**, 582–594.

Mark CA, Bushey CL, Smetanka W (1995) Fuel model identification and mapping for fire behavior prediction in the Absaroka-Beartooth Wilderness, Montana and Wyoming. In 'Proceedings: symposium on fire in wilderness and park management'. (Technical Coordinators JK Brown, RW Mutch, CW Spoon and RH Wakimoto) USDA Forest Service General Technical Report INT-GRT-320, pp. 227–229.

Miller W, Johnston D (1985) Comparison of fire fuel maps produced using MSS and AVHRR data. In 'Proceedings of the Pecora X Symposium'. pp. 305–314. (American Society for Photogrammetry and Remote Sensing: Falls Church, VA)

Millington AC, Critchley RW, Douglas TD, Ryan P (1994) 'Estimating woody biomass in sub-Saharan Africa.' (The World Bank: Washington, D.C.)

Perry GLW, Sparrow AD, Owens IF (1999) A GIS-supported model for the simulation of the spatial structure of wildland fire, Cass Basin, New Zealand. *Journal of Applied Ecology* **36**, 502–518. doi:10.1046/J.1365-2664.1999.00416.X

Pyne SJ, Andrews PL, Laven RD (1996) 'Introduction to wildland fire, 2nd edn.' (John Wiley and Sons: New York) 769 pp.

Racine CH, Dennis JG, Patterson WA (1985) Tundra fire regimes in the Novak River Watershed, Alaska (USA): 1956–1983. *Arctic* **38**, 194–200.

Rabii HA (1979) An investigation of the utility of Landsat-2 MSS data to the fire-danger rating area, and forest fuel analysis within Crater Lake National Park, Oregon. Ph.D dissertation, Oregon State University. 410 pp.

Razafimpanilo H, Frouin R, Iacobellis SF, Somerville RCJ (1995) Methodology for estimating burned area from AVHRR reflectance data. *Remote Sensing of Environment* **54**, 273–289. doi:10.1016/0034-4257(95)00154-9

Reich RM, Davis RA (1998) 'On-line spatial library for the S-PLUS© statistical software package.' (Colorado State University: Fort Collins, CO)

Roberts DA, Gardner M, Regelbrusse J, Pedreros D, Ustin S (1998) Mapping the distribution of wildland fuels using AVIRIS in the Santa Monica Mountains. *Remote Sensing of Environment* **65**, 267–279. doi:10.1016/S0034-4257(98)00037-6

Rothermel RC (1972) 'A mathematical model for predicting fire spread in wildland fuels.' USDA Forest Service, Intermountain Forest and Range Experiment Station Research Paper INT-115. 40 pp.

Salazar LA (1982) 'Remote sensing techniques aid in preattack planning for fire management.' USDA Research paper PSW-162. 19 pp.

Schloeder CA, Zimmermann NE, Jacobs MJ (2001) Comparison of methods for interpolating soil properties using limited data. *American Society of Soil Science Journal* **65**, 470–479.

Shasby MB, Burgan RE, Johnson RR (1981) Broad area forest fuels and topography mapping using digital Landsat and terrain data. In 'Proceedings of 7th international symposium on machine processing of remotely sensed data, West Lafayette', pp. 529–537.

Syrjanen K, Kalliola R, Puolasmaa A, Mattson J (1994) Landscape structure and forest dynamics in subcontinental Russian European taiga. *Annales Zoologici Fennici* **31**, 19–34.

Stone M (1974) Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B* **36**, 111–133.

Turner MG, Hargrove WW, Gardner RH, Romme WH (1994) Effects of fire on landscape heterogeneity in Yellowstone National Park. *Journal of Vegetation Science* **5**, 731–742.

Upton GJG, Fingleton B (1985) 'Spatial data analysis by example. Vol. 1, Point pattern and quantitative data.' (John Wiley and Sons: New York) 409 pp.

Williams MS (1997) A regression technique accounting for heteroscedastic and asymmetric error. *Journal of Agricultural Biological & Environmental Statistics* **2**, 108–129.

Wilson BA, Ow CFW, Heathcott M, Milne D, McCaffrey TM, Franklin SE (1994) Landsat MSS classification of fire fuel types in Wood Buffalo National Park. *Global Ecology and Biogeography Letters* **4**, 33–39.